WHAT IS CLAMED IS:

1.   A load balancing method in a system comprising a plurality of computers for processing transaction processing requests originating from a plurality of terminals, comprising steps of:

5   a) estimating load states of respective ones of the computers;

b) determining estimated elongation rates of processing time for respective ones of the computers based on the estimated load states;

10   c) calculating load indexes of respective ones of the computers from the estimated elongation rates; and

d) determining load distribution among the computers based on the load indexes.

2.   The load balancing method according to claim 1,

15   wherein the step (a) comprises the steps of:

a.1) measuring load data of each of the computers at constant intervals; and

a.2) estimating a load state of each of the computers based on the load data.

3. The load balancing method according to claim 2, wherein,

in the step (a), a current number of in-process transactions in each of the computers is measured in response
5   to start and termination of transaction processing at each of the computers, and

in the step (b), an estimated elongation rate of processing time for each of the computers is determined based on the estimated current load state and the current number of
10 in-process transactions in the computer.

4. The load balancing method according to claim 1, wherein the steps (a) through (d) are sequentially performed when a transaction processing request has been received from a terminal,
15   wherein, in the step (d), one of the computers is selected based on the load indexes as an execution computer processing the transaction processing request.

5. The load balancing method according to claim 2, wherein, in the step (a.1), a number of in-process
20 transactions and a number of job processing processes staying in a CPU system of the computer are measured as the load data measured at constant intervals.

6.   The load balancing method according to claim 2,
wherein, in the step (a.1), a number of in-process
transactions and a CPU utilization of the computer are
measured as the load data measured at constant intervals.

5   7.   The load balancing method according to claim 1,
wherein the step (a) comprises the steps of:
        a.1) measuring load data of each of the computers
at constant intervals to produce a sequence of load data; and
        a.2) estimating a current load state of each of the
10   computers based on the sequence of load data.

        8.   The load balancing method according to claim 3,
wherein the step (a) comprises the steps of:
        a.1) measuring load data of each of the computers
at constant intervals to produce a sequence of load data; and
15   a.2) estimating a current load state of each of the
computers based on the sequence of load data.

        9.   The load balancing method according to claim 3,
wherein the step (b) comprises the steps of:
        b.1) correcting the estimated load states using the
20   current numbers of in-process transactions for respective
ones of the computers to produce corrected estimated load
states; and

     b.2) determining estimated elongation rates of
processing time for respective ones of the computers based on
the corrected estimated current load states.

     10.   The load balancing method according to claim 8,
5   wherein the step (b) comprises the steps of:
     b.1) correcting the estimated load states using the
current numbers of in-process transactions for respective
ones of the computers to produce corrected estimated load
states; and
10   b.2) determining estimated elongation rates of
processing time for respective ones of the computers based on
the corrected estimated current load states.

     11.   The load balancing method according to claim 1,
wherein, in the step (c), an estimated elongation rate is
15   used as a load index of each of the computers.

     12.   The load balancing method according to claim 1,
wherein, in the step (c), an estimated elongation rate is one
of a before-scheduling estimated elongation rate and an
after-scheduling estimated elongation rate,
20   wherein the before-scheduling estimated elongation
rate is an estimated elongation rate calculated from a
corresponding estimated load state before the transaction

processing request is allocated to respective ones of the

computers, and

      the after-scheduling estimated elongation rate is

an estimated elongation rate calculated from a corresponding

5   estimated load state after the transaction processing request

is allocated to respective ones of the computers.

    13. The load balancing method according to claim 12,

wherein the step (c) comprises the steps of:

    c.1) multiplying the estimated elongation rate of

10   each of the computers by a current number of in-process

transactions in the computer to produce a total estimated

elongation rate of the computer; and

    c.2) determining the total estimated elongation

rate as a load index of the computer.

15   14. The load balancing method according to claim 12,

wherein the step (c) comprises the steps of:

    c.1) multiplying the estimated elongation rate of

each of the computers by a current number of in-process

transactions in the computer to produce a total estimated

20   elongation rate of the computer;

    c.2) calculating a total estimated elongation rate

difference between an after-scheduling total estimated

elongation rate and a before-scheduling total estimated

elongation rate of each of the computers; and

        c.3) determining the total estimated elongation

rate difference as a load index of the computer.

5   15.   A load balancing system comprising:

        a plurality of terminals, each of which originates

a transaction processing request;

        a plurality of computers, each of which processes a

plurality of transaction processing requests originating from

10   a plurality of terminals in parallel;

        a load estimator for estimating load states of

respective ones of the computers;

        a load data memory for storing the estimated load

states; and

15   an execution computer selector for selecting one of

the computers as an execution computer to be put in charge of

processing a transaction processing request based on load

indexes of respective ones of the computers, wherein the load

indexes are calculated from estimated elongation rates of

20   processing time for respective ones of the computers, wherein

the estimated elongation rates are determined based on the

estimated load states.

16. The load balancing system according to claim 15, wherein the load estimator measures load data of each of the computers at constant intervals and estimates a load state of each of the computers based on the load data.

5   17. The load balancing system according to claim 16, wherein the load estimator measures a current number of in-process transactions in each of the computers in response to start and termination of transaction processing at each of the computers; and

10   the execution computer selector determines an estimated elongation rate of processing time for each of the computers based on the estimated current load state and the current number of in-process transactions in the computer.

18. The load balancing system according to claim 15, wherein the execution computer selector is started up when a

15 transaction processing request has been received from a terminal and then selects one of the computers based on the load indexes as an execution computer processing the transaction processing request.

20   19. The load balancing system according to claim 16, wherein the load estimator measures a number of in-process transactions and a number of job processing processes staying

in a CPU system of the computer as the load data measured at
constant intervals.

20.   The load balancing system according to claim 16,
wherein the load estimator measures a number of in-process
5   transactions and a CPU utilization of the computer as the
load data measured at constant intervals.

21.   The load balancing system according to claim 15,
wherein the load estimator measures load data of each of the
computers at constant intervals to produce a sequence of load
10   data, and estimates a current load state of each of the
computers based on the sequence of load data.

22.   The load balancing system according to claim 17,
wherein the load estimator measures load data of each of the
computers at constant intervals to produce a sequence of load
15   data, and estimates a current load state of each of the
computers based on the sequence of load data.

23.   The load balancing system according to claim 17,
wherein the execution computer selector corrects the
estimated load states using the current numbers of in-process
20   transactions for respective ones of the computers to produce
corrected estimated load states, and determines estimated

elongation rates of processing time for respective ones of
the computers based on the corrected estimated current load
states.

24. The load balancing system according to claim 22,
5   wherein the execution computer selector corrects the
estimated load states using the current numbers of in-process
transactions for respective ones of the computers to produce
corrected estimated load states, and determines estimated
elongation rates of processing time for respective ones of
10 the computers based on the corrected estimated current load
states.

25. The load balancing system according to claim 15,
wherein the execution computer selector uses an estimated
elongation rate as a load index of each of the computers.

15 26. The load balancing system according to claim 15,
wherein an estimated elongation rate is one of a before-
scheduling estimated elongation rate and an after-scheduling
estimated elongation rate,
        wherein the before-scheduling estimated elongation
20 rate is an estimated elongation rate calculated from a
corresponding estimated load state before the transaction

processing request is allocated to respective ones of the
computers, and

      the after-scheduling estimated elongation rate is
an estimated elongation rate calculated from a corresponding

5   estimated load state after the transaction processing request
is allocated to respective ones of the computers.

     27.   The load balancing system according to claim 26,
wherein the execution computer selector multiplies the
estimated elongation rate of each of the computers by a

10   current number of in-process transactions in the computer to
produce a total estimated elongation rate of the computer,
and determines the total estimated elongation rate as a load
index of the computer.

     28.   The load balancing system according to claim 26,

15   wherein the execution computer selector multiplies the
estimated elongation rate of each of the computers by a
current number of in-process transactions in the computer to
produce a total estimated elongation rate of the computer,
calculates a total estimated elongation rate difference

20   between an after-scheduling total estimated elongation rate
and a before-scheduling total estimated elongation rate of
each of the computers, and determines the total estimated
elongation rate difference as a load index of the computer.

29.   The load balancing system according to claim 15,
further comprising:

a control node connected between the terminals and
the computers, wherein the control node includes the load

5   estimator, the load data memory, and the execution computer
selector.

30.   The load balancing system according to claim 15,
further comprising:

a switching device connecting the computers to each

10   other,

wherein each of the computers includes the load
estimator, the load data memory, and the execution computer
selector, wherein an estimated load state estimated at a
computer is transferred to all other computers through the

15   switching device so that the load state memory of each of the
computers stores same load state data.

31.   The load balancing system according to claim 30,
wherein the execution computer selector selects a computer to
which the execution computer selector belongs when the load

20   index of the computer is smaller than a value obtained by
multiplying a minimum load index among the computers by a
predetermined factor greater than 1.

32.   The load balancing system according to claim 15,
further comprising:

a switching device connecting the computers to each
other, the switching device including the load data memory
5   which is shared among the computers,

wherein each of the computers includes the load
estimator and the execution computer selector, wherein an
estimated load state estimated at a computer is transferred
to the switching device to be stored in the load state memory.

10   33.   The load balancing system according to claim 15,
further comprising:

an interim control node connected between the
terminals and the computers, wherein the interim control node
includes an interim execution computer selector for selecting
15   one of the computers according to a predetermined rule; and

a switching device connecting the computers to each
other,

wherein each of the computers includes the load
estimator, the load data memory, and the execution computer
20   selector, wherein an estimated load state estimated at a
computer is transferred to all other computers through the
switching device so that the load state memory of each of the
computers stores same load state data, and

        wherein the execution computer selector selects a computer to which the execution computer selector belongs when the load index of the computer is smaller than a value obtained by multiplying a minimum load index among the

5   computers by a predetermined factor greater than 1.

        34.   The load balancing system according to claim 15, further comprising:

        a switching device connecting the computers to each other, the switching device including the load data memory

10   which is shared among the computers,

        wherein each of the computers includes the load estimator and the execution computer selector, wherein an estimated load state estimated at a computer is transferred to the switching device to be stored in the load state memory,

15   and

        wherein the execution computer selector selects a computer to which the execution computer selector belongs when the load index of the computer is smaller than a value obtained by multiplying a minimum load index among the

20   computers by a predetermined factor greater than 1.

        35.   The load balancing method according to claim 1, wherein the estimated elongation rate E is a ratio of a processing time required for a job processing process to a

net processing time, wherein the estimated elongation rate E
is obtained by the following equation:

$$E = X/(X - P \cdot P), \text{ when } P < N,$$
$$= N + 1.0, \quad \text{when } P \geqq N,$$

5   where $X = N \cdot (P + 1)$, N is the number of in-process
transactions in a ~~CPU of the~~ computer and P is the number of
transaction processes in a CPU system of the computer.

36. The load balancing method according to claim 1,
wherein the estimated elongation rate E is obtained by the
10   following equation:

$$E = N(1-R)/(N(1-R)-R \cdot R)$$

where N is a number of in-process transactions in a CPU of
the computer and R is a CPU utilization.

37. A load balancing method in a system comprising a
15   plurality of computers for processing transaction processing
requests originating from a plurality of terminals,
comprising steps of:

a) estimating elongation rates of processing time
for respective ones of the computers, wherein an elongation
20   rate is a ratio of a processing time required for processing
a transaction to a net processing time which is a sum of CPU
time and an input/output time for processing the transaction;

b) calculating load indexes of respective ones of the computers based on the estimated elongation rates; and

c) selecting a destination computer from the computers based on the load indexes, wherein the destination

5   computer having a minimum one among the load indexes.

38.   A load balancing method in a system comprising a load balancing device for distributing transaction processing requests originating from a plurality of terminals to a plurality of execution computers, comprising steps of:

10   at the load balancing device,

a) receiving load data from each of the execution computers at regular intervals, the load data including a number of in-process transactions and one of a CPU utilization and a number of job processing processes staying

15   in a CPU system of the execution computer;

b) estimating load states of respective ones of the computers based on load data received from the execution computers;

c) determining estimated elongation rates of processing time for respective ones of the computers based on

20   the estimated load states;

d) calculating load indexes of respective ones of the computers from the estimated elongation rates; and

e) determining load distribution among the
computers based on the load indexes.

39.   A load balancing method in a system comprising a
plurality of computers and a plurality of terminals, wherein
5   each of the computers processes a transaction processing
requests originating from a terminal, comprising steps of:
        at each of the computers,
        a) estimating a load state of the computer based on
load data measured at regular intervals including a number of
10   in-process transactions and one of a CPU utilization and a
number of job processing processes staying in a CPU system of
the computer;
        b) storing the load state of the computer and a
load state received from another computer in a load state
15   memory;
        c) transferring the load state of the computer to
all other computers;
        d) when receiving a transaction processing request
from a terminal, determining estimated elongation rates of
20   processing time for respective ones of the computers based on
the load states stored in the load state memory;
        e) calculating load indexes of respective ones of
the computers from the estimated elongation rates;

f) determining based on the load indexes whether the transaction processing request should be processed by the computer or transferred to another computer;

g) when it is determined that the transaction

5   processing request should be transferred to another computer, determining a destination computer among the computers based on the load indexes to transfer it to the destination computer; and

h) when it is determined that the transaction

10   processing request should be processed by the computer, processing the transaction processing request.

40.   A load balancing method in a system comprising an interim load balancing device connecting a plurality of terminals and a plurality of execution computers, comprising

15   steps of:

at the interim load balancing device,

a) setting a predetermined distribution scheme;

b) when receiving a transaction processing request from a terminal, selecting an interim destination execution

20   computer from the execution computers according to the predetermined distribution scheme;

c) sending the transaction processing request to the interim destination execution computer;

at each of the execution computers,

d) estimating a load state of the execution computer based on load data measured at regular intervals including a number of in-process transactions and one of a CPU utilization and a number of job processing processes

5   staying in a CPU system of the execution computer;

e) storing the load state of the execution computer and a load state received from another execution computer in a load state memory;

f) transferring the load state of the execution

10   computer to all other execution computers;

g) when receiving the transaction processing request from the interim load balancing device, determining estimated elongation rates of processing time for respective ones of the execution computers based on the load states

15   stored in the load state memory;

h) calculating load indexes of respective ones of the execution computers from the estimated elongation rates;

i) determining based on the load indexes whether the transaction processing request should be processed by the

20   execution computer or transferred to another execution computer;

j) when it is determined that the transaction processing request should be transferred to another execution computer, determining a final destination computer among the

execution computers based on the load indexes to transfer it
to the final destination computer; and

k) when it is determined that the transaction
processing request should be processed by the execution
5   computer, processing the transaction processing request.

41. A load balancing method according to claim 40,
wherein, in the step (a), the predetermined distribution
scheme is a static distribution method such that the
terminals are previously divided into a plurality of groups
10 and an interim destination execution computer is determined
depending on which one of the groups a request originating
terminal belongs to.

42. A load balancing method according to claim 40,
wherein, in the step (a), the predetermined distribution
15 scheme is a static round distribution method such that the
execution computers are sequentially and repeatedly selected
as an interim destination execution computer in the arrival
order.

43. A load balancing method according to claim 40,
20 wherein the step (a) comprises the steps of:
a.1) receiving load data from each of the execution
computers at regular intervals, the load data including a

number of in-process transactions and one of a CPU

utilization and a number of job processing processes staying

in a CPU system of the execution computer;

     a.2) determining the predetermined distribution

5   scheme based on the load data so that a transaction

processing load is balanced among the execution computers.

     44. A recording medium storing a computer program for

instructing a computer of a load balancing device to

distribute transaction processing requests originating from a

10 plurality of terminals to a plurality of execution computers,

the computer program comprising steps of:

     a) receiving load data from each of the execution

computers at regular intervals, the load data including a

number of in-process transactions and one of a CPU

15 utilization and a number of job processing processes staying

in a CPU system of the execution computer;

     b) estimating load states of respective ones of the

computers based on load data received from the execution

computers;

20   c) determining estimated elongation rates of

processing time for respective ones of the computers based on

the estimated load states;

     d) calculating load indexes of respective ones of

the computers from the estimated elongation rates; and

e) determining load distribution among the
computers based on the load indexes.

45. A recording medium storing a computer program for
instructing a computer to balance a transaction processing
5   load among a plurality of computers, the computer program
comprising steps of:

a) estimating a load state of the computer based on
load data measured at regular intervals including a number of
in-process transactions and one of a CPU utilization and a
10   number of job processing processes staying in a CPU system of
the computer;

b) storing the load state of the computer and a
load state received from another computer in a load state
memory;

15   c) transferring the load state of the computer to
all other computers;

d) when receiving a transaction processing request
from a terminal, determining estimated elongation rates of
processing time for respective ones of the computers based on
20   the load states stored in the load state memory;

e) calculating load indexes of respective ones of
the computers from the estimated elongation rates;

f) determining based on the load indexes whether
the transaction processing request should be processed by the
computer or transferred to another computer;

g) when it is determined that the transaction

5   processing request should be transferred to another computer,
determining a destination computer among the computers based
on the load indexes to transfer it to the destination
computer; and

h) when it is determined that the transaction

10 processing request should be processed by the computer,
processing the transaction processing request.

46. A recording medium storing:

a first computer program for instructing a first
computer of an interim load balancing device to distribute

15 transaction processing requests originating from a plurality
of terminals to a plurality of execution computers; and

a second computer program for instructing each of the
execution computers to balance a transaction processing load
among the execution computers,

20   wherein the first computer program comprises the
steps of:

a) setting a predetermined distribution scheme;

b) when receiving a transaction processing request from a terminal, determining an interim execution computer according to the predetermined distribution scheme; and

c) sending the transaction processing request to

5   the interim execution computer,

wherein the second computer program comprises the steps of:

d) estimating a load state of the execution computer based on load data measured at regular intervals

10   including a number of in-process transactions and one of a CPU utilization and a number of job processing processes staying in a CPU system of the execution computer;

e) storing the load state of the execution computer and a load state received from another execution computer in

15   a load state memory;

f) transferring the load state of the execution computer to all other execution computers;

g) when receiving the transaction processing request from the interim load balancing device, determining

20   estimated elongation rates of processing time for respective ones of the execution computers based on the load states stored in the load state memory;

h) calculating load indexes of respective ones of the execution computers from the estimated elongation rates;

FQ5-489   81

i) determining based on the load indexes whether the transaction processing request should be processed by the execution computer or transferred to another execution computer;

5   j) when it is determined that the transaction processing request should be transferred to another execution computer, determining a final destination computer among the execution computers based on the load indexes to transfer it to the final destination computer; and

10   k) when it is determined that the transaction processing request should be processed by the execution computer, processing the transaction processing request.